Graph Adversarial Self-Supervised Learning (Appendix)

Anonymous Author(s)
Affiliation
Address
email

1 Related work

1.1 Semi-supervised learning

In the semi-supervised setting [1, 2], it is common to reduce over-fitting by combining classification loss with an unsupervised loss. A commonly used teacher-student network consists of a *teacher* network and a *student* network, where the former generates targets for the latter. Models evaluates the consistency loss of predictions of data with and without noise [3], the outputs in different passes [4], the exponential moving average (EMA) prediction for each training samples [4], or the EMA of weights [5]. Unlike the above methods that use random noise, [6] directly learn a tiny perturbation in adversarial direction to add to the input which would most significantly affect the output of the prediction function. The above mentioned method are effective in semi-supervised learning case for image classification task, our GASSL follows the structure of BYOL [7] and differs from MT [5] in having an extra predictor on the online network to prevent collapse and focuses on how to learn effective representations of graphs without human supervision.

2 Experimental Setup

2.1 Datasets

2.1.1 TU datasets

MUTAG [8] is a dataset of 188 mutagenic aromatic and heteroaromatic nitro compounds. PTC [8] is a dataset of 344 chemical compounds that reports the carcinogenicity for male and female rats. NCI1 [8] is a subset of balanced datasets of chemical compounds screened for the ability to suppress or inhibit the growth of a panel of human tumor cell lines. IMDB-BINARY [9] and its multi-class version IMDB-MULTI [9] are movie-collaboration datasets collected actor/actress and genre information of different movies on IMDB. COLLAB [9] is a scientific-collaboration dataset, derived from 3 public collaboration datasets.

2.1.2 OGB datasets

HIV is adopted from the MoleculeNet [10], and is among the largest of the MoleculeNet datasets. Tox21, ToxCast, and BBBP are three smaller datasets from MoleculeNet. All the molecules are pre-processed using RDKit. Each graph represents a molecule, where nodes are atoms, and edges are chemical bonds. Input node features are 9-dimensional, containing atomic number and chirality, as well as other additional atom features such as formal charge and whether the atom is in the ring or not. We use the features provided by OGB and more detailed information can be found at `https://ogb.stanford.edu/docs/graphprop/`.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

Table 1: Effect of ascent step T on the accuracy on graph classification task.

Encoder	T	MUTAG	PTC-MR	IMDB-B	IMDB-M	COLLAB	NCI1
GCN	1	89.3 ± 6.0	58.9 ± 7.9	$\mathbf{73.0 \pm 0.3}$	49.2 ± 3.1	77.8 ± 1.9	76.7 ± 2.1
	2	89.4 ± 6.9	60.5 ± 6.8	72.4 ± 0.7	49.1 ± 1.9	$\mathbf{77.9 \pm 2.1}$	76.8 ± 3.0
	3	$\mathbf{90.4 \pm 7.9}$	$\mathbf{62.2 \pm 6.0}$	72.7 ± 0.7	$\mathbf{49.6 \pm 2.3}$	$\mathbf{77.9 \pm 2.0}$	$\mathbf{77.0 \pm 1.9}$
GIN	1	89.8 ± 5.8	63.6 ± 5.3	73.0 ± 0.3	51.5 ± 2.3	$\mathbf{78.0 \pm 2.0}$	80.1 ± 2.2
	2	89.3 ± 7.1	62.8 ± 5.8	72.8 ± 0.5	51.2 ± 2.2	77.5 ± 1.8	$\mathbf{80.2 \pm 2.1}$
	3	$\mathbf{90.9 \pm 7.9}$	$\mathbf{64.6 \pm 6.1}$	$\mathbf{74.2 \pm 0.5}$	$\mathbf{51.7 \pm 2.5}$	$\mathbf{78.0 \pm 2.0}$	$\mathbf{80.2 \pm 1.9}$

Table 2: Effect of ascent steps T on the accuracy and training cost (in seconds) for 200 epochs on IMDB-BINARY, COLLAB, and NCI1 datasets.

Encoder	T	IMDB-B	Speed-up	COLLAB	Speed-up	NCI1	Speed-up
GIN	1	170	1x	960	1x	357	1x
	2	168	1.01x	948	1.01x	340	1.05x
	3	150	1.13x	926	1.04x	302	1.18x

2.2 Linear evaluation

We use the Graph Isomorphism Network (GIN) [11] and Graph Convolutional Network (GCN) [12] as encoder in our GASSL. GNN layers are chosen from $\{2, 3, 4, 5\}$. The number of epochs are chosen from $\{20, 40, 100, 200\}$.

For TU datasets, we follow [13] and use LinearSVC as linear classifier for linear evaluation. We split the data into training data and test data by $9:1$ with StratifiedKFold. The parameter C is chosen from $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$. The best parameter is chosen by grid search with 5-fold cross validation on training data.

For OGB datasets, we follow [14] and attach the logistic regression on the top of the frozen GNN encoder for linear evaluation. We split the training data, validation data and test data in the ratio of $8:1:1$. We train the linear classifier within 100 epochs. We evaluate the classification accuracy on the validation set per 10 epochs to obtain the best parameters. We train the linear classifier using Adam optimizer with an initial learning rate of $3e-4$.

3 Additional Experiment Results

3.1 Effect of ascent steps T

We explored the impact of ascent step T on the graph classification accuracy of six TU dataset. We examined both GCN and GIN encoders and set the ascent step $T \in \{1, 2, 3\}$. From table (1), we observed that our method achieves a stable performance that essentially does not change as the ascent step changes. In particular, we note that `GASSL` performs best at $T = 3$. With the gradient accumulation strategy, we achieve training speedup on multiple datasets. For example, from Table 2 there is a 1.18x speedup on the NCI1 dataset. We reduce the training time by reducing the number of outer loops. However, since the adversarial training needs to solve the minimum and maximum problems alternately, it is still a direction for our future research in improving the training efficiency.

References

[1] Olivier Chapelle, Bernhard Schlkopf, and Alexander Zien. *Semi-Supervised Learning*. 2010.

[2] Xiaojin Zhu, Andrew B. Goldberg, Ronald Brachman, and Thomas Dietterich. *Introduction to Semi-Supervised Learning*. 2009.

[3] Antti Rasmus, Harri Valpola, Mikko Honkala, Mathias Berglund, and Tapani Raiko. Semi-supervised learning with ladder networks. In *NIPS'15 Proceedings of the 28th International*

Conference on Neural Information Processing Systems - Volume 2, volume 28, pages 3546–3554, 2015.

[4] Samuli Matias Laine and Timo Oskari Aila. Temporal ensembling for semi-supervised learning, 2017.

[5] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, volume 30, pages 1195–1204, 2017.

[6] Takeru Miyato, Andrew M. Dai, and Ian J. Goodfellow. Adversarial training methods for semi-supervised text classification. In *ICLR (Poster)*, 2017.

[7] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[8] Nils Kriege and Petra Mutzel. Subgraph matching kernels for attributed graphs. In *Proceedings of the 17th International Conference on Machine Learning (ICML 2012)*. Morgan Kaufmann, 2012.

[9] Pinar Yanardag and S.V.N. Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, page 1365–1374, New York, NY, USA, 2015. Association for Computing Machinery.

[10] Nikil Wale and George Karypis. Comparison of descriptor spaces for chemical compound retrieval and classification. In *Sixth International Conference on Data Mining (ICDM'06)*, volume 14, pages 347–375, 2006.

[11] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks. In *International Conference on Learning Representations*, 2018.

[12] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

[13] Fan-Yun Sun, Jordan Hoffman, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *International Conference on Learning Representations*, 2019.

[14] Minki Kang Jinheon Baek and Sung Ju Hwang. Accurate learning of graph representations with graph multiset pooling. In *ICLR 2021*, 2021.